

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079242 A2(51) International Patent Classification: **C07K 14/195**(21) International Application Number: **PCT/IB02/02080**(22) International Filing Date: **28 March 2002 (28.03.2002)**(25) Filing Language: **English**(26) Publication Language: **English**(30) Priority Data:
0108024.1 30 March 2001 (30.03.2001) **GB**(71) Applicant (for all designated States except US): **CHIRON SPA** [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MASIGNANI, Vega** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT). **PIZZA, Mariagrazia** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT). **RAPPUOLI, Rino** [IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT).(74) Agents: **HALLYBONE, Huw, George et al.**; Carpmac's & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IE, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **ADP-RIBOSYLATING BACTERIAL TOXINS**

<i>E.coli</i> LT	KLYRADSRPPD	LYDHARGTQ	YDDGYVSTSLSLRSA	SPHPYEQEVSA LGGI
<i>N.meningitidis</i>	FLYRGISCQQD	VYAHQIETG	YDGCYISTTTDKETIA	PENPNEKEVTIRAED
<i>S.coelicolor</i>	TLYRSDSRGPQVVFEEGFHAKDVQNGQYDVEKYVLVNQPSPIVSTSYDHDLY			HWADQVEVAFPGGI
<i>M.pneumoniae</i>	FVYRVDLRSPE	FFEHLSTN	GRSYFISTSETPTAA	TSFAYQREWFDTGPI
<i>S.typhi</i>	FVYRV DSTPPD		SCSGGSSDSRYIATT	TMRLQREYVSTLSI
<i>S.paratyphi</i>	FVYRV DSTPPD		SCSGGSSDSRYIATT	TMRLQREYVSTLSI
<i>S.pyogenes</i>	VVYRYVYETFL		TKHSFMSTTALKNGA	SAVPSEVELLFPRGC

(57) Abstract: ADP ribosylating toxins from *Neisseria meningitidis*, *Streptomyces coelicolor*, *Mycoplasma pneumoniae*, *Salmonella typhimurium*, *Salmonella paratyphi*, and *Streptococcus pyogenes* are disclosed, together with mutant toxins and uses therefor. There is only a low level of sequence identity between these toxins and toxins such as cholera toxin and *E.coli* heat labile toxin.

WO 02/079242 A2

ADP-RIBOSYLATING BACTERIAL TOXINS

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention is in the field of ADP-ribosylating bacterial toxins and their uses.

5 BACKGROUND ART

ADP-ribosylating bacterial exotoxins are widely known. Examples include diphtheria toxin (*Corynebacterium diphtheriae*), exotoxin A (*Pseudomonas aeruginosa*), cholera toxin (CT; *Vibrio cholerae*), heat-labile enterotoxin (LT; *E.coli*) and pertussis toxin (PT).

10 The toxins catalyse the transfer of an ADP-ribose unit from NAD^+ to a target protein. CT, for instance, transfers ADP-ribose to a specific arginine side chain of the α subunit of G_s , which blocks the ability of G_s to hydrolyse GTP to GDP. This locks the protein in its 'active' form, so adenylate cyclase activity is permanently activated. Cellular cAMP levels rise, leading to the active transport of ions from the cell and the loss of water into the gut [1].

15 The toxins are typically divided into two functionally distinct domains – A and B. The A subunit is responsible for the toxic enzymatic activity, whereas the B subunit is responsible for cellular binding. The subunits might be domains on the same polypeptide chain, or might be separate polypeptide chains. The subunits may themselves be oligomers *e.g.* the A subunit of CT consists of A_1 and A_2 which are linked by a disulphide bond, and its B subunit is a homopentamer. Typically, initial contact with a target cell is mediated by the B subunit and then subunit A alone enters the cell.

20 Crystal structures [2] are known for LT [3], CT [4] and PT [5].

The toxins are typically immunogenic, and have been proposed for use in acellular vaccines. One problem, however, is that the proteins retain their toxic activity in the vaccines. To avoid this problem, site-directed mutagenesis of key active site residues has been used to remove toxic enzymatic activity whilst retaining immunogenicity [*e.g.* refs. 6 (CT and LT), 7 (PT), 8 *etc.*]. Current
25 acellular whooping cough vaccines include a form of pertussis toxin with two amino acid substitutions ($\text{Arg}^9 \rightarrow \text{Lys}$ and $\text{Glu}^{129} \rightarrow \text{Gly}$; 'PT-9K/129G' [9]).

As well as their immunogenic properties, the toxins have been used as adjuvants. Parenteral adjuvanticity was first observed in 1972 [10] and mucosal adjuvanticity in 1984 [11]. It was surprisingly found in 1993 that the detoxified forms of the toxins retain adjuvanticity [12].

30 It is an object of the invention to provide further ADP-ribosylating bacterial toxins.

DISCLOSURE OF THE INVENTION

The amino acid sequences of six different ADP-ribosylating toxins from Gram positive and Gram negative bacteria are given in SEQ IDs 1, 3, 4, 5, 6 and 7. These toxins are from *Neisseria meningitidis*, *Streptomyces coelicolor*, *Mycoplasma pneumoniae*, *Salmonella typhimurium*,

Salmonella paratyphi, and *Streptococcus pyogenes*. The existence of ADP-ribosylating toxins in these bacterial species has not previously been suggested and, furthermore, there is only a low level of sequence identity between these toxins and toxins such as CT, LT and PT.

Toxins of the invention

- 5 The invention provides a protein having an amino acid sequence consisting of one of SEQ IDs 1, 3, 4, 5, 6 and 7. These proteins are referred to hereafter as "toxins of the invention".

Mutant toxins of the invention

- The invention also provides a protein having an amino acid sequence consisting of one of SEQ IDs 1, 3, 4, 5, 6 and 7, except that the amino acid sequence contains one or more mutations. The mutation(s)
10 preferably reduce or eliminate the ADP-ribosylating activity of the protein.

The mutations may each independently be a substitution, an insertion, or a deletion. Preferably, the amino acid sequences contains fewer than twenty mutations (*e.g.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1). Each mutation preferably involves a single amino acid.

Preferred mutations are single amino acid substitutions (*e.g.* SEQ IDs 10, 12, 14 & 16).

- 15 The invention also provides a process for diminishing the ADP-ribosylating enzymatic activity of a toxin of the invention, comprising mutating one or more amino acid residues of said toxin. This may conveniently be achieved by performing site-directed mutagenesis on nucleic acid encoding the toxin. The invention further provides a protein obtainable by this process.

Mutations may also be introduced to improve stability *e.g.* the insertion of disulphide bonds [13].

- 20 The proteins defined above are referred to hereafter as "mutant toxins of the invention" or "toxoids of the invention".

Preferred sites for mutation are given in Table 1, together with preferred mutations at those sites.

Proteins of the invention

- The invention provides a protein comprising the amino acid sequence of a toxin or mutant toxin of
25 the invention.

- It also provides a protein comprising an amino acid sequence having sequence identity to the amino acid sequence of a toxin or mutant toxin of the invention. The degree of sequence identity is preferably greater than 50% (*e.g.* 60%, 70%, 80%, 90%, 95%, 99% or more). These proteins include homologs, orthologs, allelic variants and functional mutants. Typically, 50% identity or more
30 between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as

implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides a protein comprising a fragment of a toxin or mutant toxin of the invention. The fragment should comprise at least *n* consecutive amino acids from the toxin or mutant toxin and, depending on the particular sequence, *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more). Preferably the fragment comprises one or more epitopes from the toxin. Preferred fragments are those common to two or more of SEQ ID¹ 1, 3, 4, 5, 6 and 7.

The toxins, mutant toxins and proteins defined above are collectively referred to hereafter as the "proteins of the invention".

The proteins of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from native host, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from host cell proteins).

The invention also provides the proteins of the invention (particularly the mutant toxins) for use as adjuvants and, in particular, as mucosal adjuvants.

The invention also provides the use of proteins of the invention in the manufacture of a medicament for raising an immune response in an animal. The medicament is preferably an immunogenic compositions (*e.g.* a vaccine), and will comprise, in addition to a protein of the invention, an antigen against which an immune response is to be raised. The medicament is preferably administered mucosally *e.g.* orally or intranasally.

The invention also provides immunogenic compositions (*e.g.* a vaccine) comprising a protein of the invention in admixture with a second antigen. It also provides a kit comprising a protein of the invention and a second antigen for simultaneous, separate or sequential administration. The second antigen is preferably one of the *N.meningitidis* proteins disclosed in references 14 to 20. The composition may comprise a third antigen, a fourth antigen, a fifth antigen *etc.*, one or more of which may be selected from the *N.meningitidis* proteins disclosed in these seven references.

According to a further aspect, the invention provides antibody which binds to a protein of the invention. These may be polyclonal or monoclonal and may be produced by any suitable means. The antibody may include a detectable label.

According to a further aspect, the invention provides nucleic acid encoding the proteins of the invention. Nucleic acid comprising a fragment of these coding sequences are also provided. These should comprise at least *n* consecutive nucleotides from the coding sequence and, depending on the particular sequence, *n* is 10 or more (*e.g.* 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These can be used in hybridisation reactions (*e.g.* Northern or Southern blots, or

in nucleic acid microarrays or 'gene chips') and amplification reactions (*e.g.* PCR, SDA, SSSR, LCR, NASBA, TMA) *etc.*

The invention also provides nucleic acid comprising one or more of SEQ IDs 2, 8, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25 and/or 26.

- 5 It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*e.g.* for antisense or probing, or for use as primers).

Nucleic acid according to the invention can, of course, be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, primers, probes *etc.*).

- 10 Nucleic acid according to the invention may be labelled *e.g.* with a radioactive or fluorescent label. This is particularly useful where the nucleic acid is to be used as a primer or probe *e.g.* in PCR, LCR or TMA.

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

- 15 According to a further aspect, the invention provides vectors comprising nucleic acid of the invention (*e.g.* cloning or expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as immunogenic compositions, for instance, or as diagnostic reagents, or as vaccines.

- 20 The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*e.g.* vaccines) or diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody of the invention in the manufacture of: (i) a medicament for treating or preventing bacterial infection; (ii) a diagnostic reagent for detecting the presence of bacteria or of antibodies raised against bacteria; and/or (iii) a reagent which can raise antibodies against bacteria. Said bacteria
25 are preferably *Neisseria meningitidis*, *Streptomyces coelicolor*, *Mycoplasma pneumoniae*, *Salmonella typhimurium*, *Salmonella paratyphi*, or *Streptococcus pyogenes*.

The invention also provides a method of treating a patient, comprising administering a therapeutically effective amount of nucleic acid, protein, and/or antibody of the invention.

- 30 The invention also provides a kit comprising primers (*e.g.* PCR primers) for amplifying a target sequence contained within a bacterium nucleic acid sequence, the kit comprising a first primer and a second primer, wherein the first primer is substantially complementary to said target sequence and the second primer is substantially complementary to a complement of said target sequence, wherein

the parts of said primers which have substantial complementarity define the termini of the target sequence to be amplified.

According to further aspects, the invention provides various processes.

5 A process for producing a protein of the invention is provided, comprising the step of culturing a host cell of to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

10 A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridising conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting bacteria in a biological sample is also provided, comprising the step of contacting nucleic according to the invention with the biological sample under hybridising conditions. The process may involve nucleic acid amplification (*e.g.* PCR, SDA, SSSR, LCR, TMA *etc.*) or hybridisation (*e.g.* microarrays, blots, hybridisation with probe in solution *etc.*).

15 A process for detecting a protein of the invention is provided, comprising the steps of: (a) contacting an antibody of the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

20 A summary of standard techniques and procedures which may be employed to perform the invention (*e.g.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual*, 25 *Third Edition* (2001); *DNA Cloning, Volumes I and II* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for* 30 *Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

- 5 The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a
10 gene is operably linked to the gene. Another example is where a staphylococcus sequence is heterologous to a mouse host cell. A further example would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide
15 replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

20 As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic
25 variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (*eg.* see US patent 5,753,235).

Expression systems

Nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

30 i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream
35 of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100-200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook *et al.*].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes (e.g., the murine metallothionein gene) also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, as they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *PNAS USA* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot & Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor

and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers & Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers & Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This may

contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extra-chromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32bp downstream from the ATT; see Luckow & Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers & Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers & Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers & Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, etc. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also present in the medium, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in Jones & MacMillan, pages 21-52 of *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel & Hickey, *PNAS USA*, 84:1337-1339 (1987).

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink & Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed & Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *PNAS USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *PNAS USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Herero-callis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary between plant species, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on genotype, and on the history of the culture. If these three variables are controlled then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *E.coli* [Raibaud *et al.* (1984) *Annu.Rev.Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc.Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl.Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The β -lactamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *PNAS USA* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *PNAS USA* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EP-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E.coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the

junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

- 5 Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the
10 inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

- DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*)
15 [Oka *et al.* (1985) *PNAS USA* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *PNAS USA* 79:5582; EP-A-0 244 042].

- Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct
20 the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

- Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence
25 of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy
30 number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

- Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector.
35 Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469].

Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *PNAS USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *PNAS USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *PNAS USA* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *PNAS USA* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColEI-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. Boyer & Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, *Streptococcus*].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the

absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, so sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *PNAS USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *PNAS USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and

the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase (EP-A-0012873; JPO 62,096,086) and A-factor (US patent 4,588,684) genes. Alternatively, leaders of
5 non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0060057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008;
10 EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (eg. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct
15 the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression
20 constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *PNAS USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.*
25 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually ~10 to ~150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least ~20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be
35 directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *PNAS USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector,

or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *PNAS USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *PNAS USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid

antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying proteins of the invention.

- 5 Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete
10 adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the
15 immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (*eg.* 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

- Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However,
20 rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to
25 fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*eg.* hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (*eg.* in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

- 30 If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ³²P and ¹²⁵I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific
35 binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ¹²⁵I may serve as a radioactive label or as

an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon their size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

5 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

10 Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 etc.), interferons (eg. γ interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, their taxonomic group (eg. nonhuman primate, primate, etc.), the capacity of their immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, eg. by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (eg. WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be used [eg. Robinson & Torres (1997) *Seminars in Immunol* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically.

These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *PNAS USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769,

WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992)

5 *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10
10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native
15 D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935,
20 and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

25 The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary HSV vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human*
30 *Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC with accession numbers VR-977 and VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in
35 US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques.
40 Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *PNAS USA* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *PNAS USA* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see USSN 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eukaryotic cell delivery vehicles cells, for example see USSN 08/240,030, filed May 9, 1994, and USSN 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *PNAS USA* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations

like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *PNAS USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *PNAS USA* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. Subjects to be treated can be birds or mammals (including humans).

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene-mediated transfection, protoplast fusion, electroporation, encapsulation of polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

One example are polypeptides which include, without limitation: asialoglycosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *PNAS USA* 84:7413-7416); mRNA (Malone (1989) *PNAS USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *PNAS USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *PNAS USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *PNAS USA* 76:3348; Enoch & Strittmatter (1979) *PNAS USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *PNAS USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C & E, over time these lipoproteins lose A and acquire C & E. VLDL comprises A, B, C & E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *PNAS USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in

conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, MA, USA. Further description of lipoproteins can be found in WO98/06437..

10 F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, *etc.*

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/EBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Proteins of the invention can be used in immunoassays to detect antibody levels (or, conversely, antibodies of the invention can be used to detect protein levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules.

Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\% \text{ formamide}) - 600/n - 1.5(\% \text{ mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are non-homologous, and as a result, background decreases. If a radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of washes affects the intensity of the hybridizing band and the degree of background in a similar way. The stringency of washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex that is stable enough to be detected. Nucleic acid probes will hybridize to the nucleic acid of the invention (sense and/or antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the wild-type sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to a sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a bacterial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a bacterial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions (*e.g.* temperature, salt condition *etc.*). For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*PNAS USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* Agrawal & Iyer (1995) *Curr.Opin. Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as PNAs may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acid. The assay is described in Mullis *et al.* [*Meth. Enzymol.* (1987) 155:335-350] & US patents 4,683,195 & 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired bacterial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the bacterial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows an alignment of catalytic domains of various bacterial toxins, including the *N.meningitidis* toxin of the invention (NMB1343). Residues important for catalytic activity are shown enlarged.

Figure 2 shows a multiple sequence alignment of conserved regions of LT and toxins of the invention. Residues important for catalytic activity are underlined. Residues important for the conservation of structure have a shaded background. Other conserved residues are indicated in bold.

Figure 3 shows the incorporation of radio-labelled NAD into the *N.meningitidis* toxin. In Figure 3B, the lanes are: (L1) 95°C; (L2) Novobiocin, 5 mMol; (L3) GTP, 10 mMol; (L4) ATP, 10 mMol; (L5) ADP-ribose, 10 mMol; (L6) Nicotinamide (NAM), 10 mMol; (L7) control. The arrow shows the position of the toxin.

A protein with amino acid sequence <SEQ ID 1> was identified in *N.meningitidis* (serogroup B, strain MC58). This is encoded by a gene having nucleotide sequence <SEQ ID 2>.

```

1343          MGNFLYRGISCQODE-QNNGQLKPKGNKAEVAIRYDGKFKYDGKATHGPSVKNAV
          |||:| ||::| | |:|:: :| :| :| :::
CT-A  MVKIIIFVFFIFLSSFSYANDDKLYRADSRPPDEIKQSGGLMPRGQS-----EY-----FD---RGTQMNIINL

1343  YAHQ--IETGL--YDGCYISTTTDKEIAKKFATS--SGIENGYIYVLNR--DLFGQYSIFEYEVEHPENPNK
| | :||: :| |:|:: : : :| : : || : ||||: ::|: :: : ||: |:
CT-A  YDHARGTQTGFVRHDDGYVSTSISLRS AHLV GQTILSGHSTYYIYVIATAPNMFNVNDVLGAYSPHPD---EQ

1343  EVTIRAEDCGCIPEEVI IAKELIEIN*
| |:
CT-A  EVSALGGIPYSQIYGWYRVHFGVLDEQLHRNRGYRDRYYSNLDIAPAADGYGLAGFPPEHRAWREEPWIIHAP

CT-A  PGCGNAPRSSMSNTCDEKTS LGVKFLDEYQSKVKRQIFSGYQSDIDTHNRIKDEL

```

```

1343          MGNFLYRGISCCQDE-QNNGQLKPKGNKAEVAIRYDGFYDGGKATHGPSVKNAV
          | : ||| : |  || : : | | :| : :      : |      : | : : :
LT-A  MKNITFIFFILLASPLYANGDRLYRADSRPPDEIKRSGGLMPRGHNE---YFD-----RGTQMNINL

1343  YAHQ--IETGL--YDGCYISTTTDKEIAKKFATS--SGIENGYIYVLNR-----DLFGQYSIFEYEVEHP
          | |      :||: || |:||: : : | |  | | : ||||:      | :| || : || |
LT-A  YDHARGTQTGFVRYDDGYVSTLSLSLRSAGLAGQSILSGYSTYYIYVIATAPNMFNVNDVLGVYSPHPYEQEV

1343  ENPNEKEVTIRAEDCGCIPEEVIIAKELIEIN*

          ..
LT-A  ALGGIPYSQIYGWYRVNFGVIADERLHRNREYRDRYYRNLNIAPAEDGYRLAGFPDPHQAWREEPWIHHAPQG

LT-A  CGNSSRTITGDTCNEETONLSTIYLREYQSKVKRQIFSDYQSEVDIYNRIRDEL258

```

No corresponding gene was identified in *N.meningitidis* serogroup A (strain Z2491 [21]) or in *N.gonorrhoeae*.

A protein with amino acid sequence <SEQ ID 3> [CAB76015] was identified in *S.coelicolor*:

This protein shows 27% identity to CT-A over 136aa overlap:

10

15

20

25

As for *N.meningitidis*, therefore, there is only a low level of overall identity with known toxins, but key catalytic residues are conserved. The database annotation of ‘putative secreted protein’ does not suggest ADP ribosyltransferase activity.

30

35

45

50

55

60

seqid4 QYASSVKELEDTPVYLRGIKTQKTFMLQADPQNNVFLVEVNPKOKSSFOTIFFWDVYO

As for *N.meningitidis* and *S.coelicolor*, therefore, there is only a low level of overall identity with a known toxin, but key catalytic residues are conserved. The database annotation of ‘hypothetical protein’ does not suggest ADP ribosyltransferase activity.

A protein with the amino acid sequence <SEQ ID 5> was identified in *S.typhimurium*:

```

1  MKKLIFLTLS  IVSFNNYAVD  FVYRV DSTPP  DVIFRDGFSL  LGYNRNFQQF
51  ISGRSCSGGS  SDSRYIATTS  SVNQTYAIAR  AYYSRSTFKG  NLYRYQIRAD
101 NNFYSLLPSI  TYLETQGGHF  NAYEKTMMRL  QREYVSTLSI  LPENIQKAVA
151 LVYDSATGLV  KDGVTSMNNS  YLGLSTSNP  GVIPFLPEQP  TYTQRIXAF
201 GPLISSCSFI  GSVCHSHRQG  RADVYNMSFY  DARPVLEIL  SK

```

SEQID5 MKKLI²⁰FL²¹TL²²SIV²³SF²⁴NN²⁵YAV²⁶DF²⁷VY²⁸RV²⁹DST³⁰PPD³¹VIF³²RD³³GS³⁴LL³⁵GY³⁶NR³⁷NF³⁸QQ³⁹FI⁴⁰
 ||| ||| :: :||: ||| : :
 pt . pep ARTGWL²⁰TW²¹LAIL²²AV²³TAP²⁴VT²⁵SPA²⁶WAD²⁷DP²⁸PAT²⁹VY³⁰RY³¹DS³²RP³³PE³⁴DF³⁵QNG³⁶FT³⁷AW³⁸GN³⁹DN⁴⁰VL⁴¹DHL⁴²
 20 30 40 50 60 70

```

SEQID5      SGRSCSGGSSDSRYIATTSS-----VNQTYAIARAYYSRSTFKGNLYRYQIRAD
             :|||: |||: ::|||               :::   ||  : : | | | :|||
pt.pep      TGRSCQVGSSNSAFVSTSSRRYTEVYLEHRMQEAVEAERAGRGTGHFIG--YIYEVRA

```

```

SEQID5      MNFYSLLPs-ITYLETQGGHFN-AYEKTMMRLQREYVSTLSILPENIQKAVALVYDSATG
            ||||:  | : |::| | : :      :  |||::  | |||:::  : ::  ||
pt.pep      MNFYGAASSYFEYVDTYGDNAGRILAGALATYQSEYLAHRRIPPENIRRVTRVYHNGITG

```

```

SEQID5      LVKDGVTMTNASYLGLSTTSNPGVIFFLPEPQTYTQQRIXAFGPLISSCFSIGSVCHSHR
:           : || |:: :| :||:      || :      ||:      ||: |::
pt.pep      ETTTTEYS-NARYVSQOTRANPN-----PYTSRRRSVASIVGTLVRMAPVIGA-CMARQ

```

```
SEQID5      GQRADVYNMSFYDARPVIELILSK
:: :: | : :: |      ::|
pt . pep    AESSEA--MAAWSERAGEAMVLVYYESIAYSF
```

As described above, therefore, there is only a low level of overall identity with a known toxin, but key catalytic residues are conserved.

In addition, a protein (SEQ ID 9) upstream gene in *S.typhimurium* shows homology to the pertussis toxin S2 subunit:

Score = 34.0 bits (77), Expect = 0.91

Identities = 31/101 (30%), Positives = 42/101 (40%), Gaps = 7/101 (6%)

Query: 30 TNAYYSDEVISLHVGQIDTSPYFCIKTVKANGSGTPVV-ACAVSKQSIWAPSKELLDQ 88
T+ YYS+ + L T+ C V+ SG PV+ AC + + L

Sbjct: 98 TDHYYSNVTATRLLS--STNSRLCAVFVR--SGQPVIGACTSPYDGKYWSMYSRLRKM 151

Query: 89 ARYFYSTGQSVRIHVQKNIWTYPLFVNTFSANALVGLSSCS 129

Y G SVR+HV K Y TF AL G+S C+
Sbjct: 152 LYLIYVAGISVRVHVSKEEOYYDYEDATFETYALTGISICN 192

Toxin gene from *Salmonella paratyphi A* (strain ATCC 9150)

A protein with amino acid sequence <SEQ ID 6> was identified in *S. paratyphi*. This shows good homology to the *S. typhimurium* sequence shown above:

```

5      Score = 1231 (438.4 bits), Expect = 1.6e-125, P = 1.6e-125
      Identities = 241/242 (99%), Positives = 241/242 (99%)

Typhi:      1 MKKLIFLTLSIVSFNNYAVDFVYRV DSTPPDVIFRDGFSLLGYNRNFQQFISGRSCSGGS 60
              MKKLIFLTLSIVSFNNYAVDFVYRV DSTPPDVIFRDGFSLLGYNRNFQQFISGRSCSGGS
Parat:      1 MKKLIFLTLSIVSFNNYAVDFVYRV DSTPPDVIFRDGFSLLGYNRNFQQFISGRSCSGGS 60

10     Query:   61 SDSRYIATTSSVNQTYAIARAYYSRSTFKGNLYRYQIRADNNFYSL LPSITYLETQGGHF 120
              SDSRYIATTSSVNQTYAIARAYYSRSTFKGNLYRYQIRADNNFYSL LPSITYLETQGGHF
Sbjct:      61 SDSRYIATTSSVNQTYAIARAYYSRSTFKGNLYRYQIRADNNFYSL LPSITYLETQGGHF 120

15     Query:   121 NAYEKTMMRLQREYVSTLSILPENIQKAVALVYDSATGLVKDGVSTMNASYLGLSTTSNP 180
              NAYEKTMMRLQREYVSTLSILPENIQKAVALVYDSATGLVKDGVSTMNASYLGLSTTSNP
Sbjct:      121 NAYEKTMMRLQREYVSTLSILPENIQKAVALVYDSATGLVKDGVSTMNASYLGLSTTSNP 180

20     Query:   181 GVIPFLPEPQTYTQQRIFA GFLISSCFSIGSVCHSHRGQRADVYNMSFYDARPVIELIL 240
              GVIPFLPEPQTYTQQRIFA GFLISSCFSIGSVCHSHRGQRADVYNMSFYDARPVIELIL
Sbjct:      181 GVIPFLPEPQTYTQQRIFA GFLISSCFSIGSVCHSHRGQRADVYNMSFYDARPVIELIL 240

      Query:    241 SK* 242
              SK*
25     Sbjct:   241 SK* 242

```

Again, this protein shows only a low level of overall identity with a known toxin, but key catalytic residues are conserved.

As for *S. typhimurium*, there is an upstream PT-S2 homolog, but it is frame-shifted:

```

30     Score = 387 (141.3 bits), Expect = 4.9e-36, P = 4.9e-36
      Identities = 73/73 (100%), Positives = 73/73 (100%), Frame = +3

Query:      65 TPVVACAVSKQSIWAPSFKELLDQARYFYSTGQSVRIHVQKNIW TYPLFVNTFSANALVG 124
              TPVVACAVSKQSIWAPSFKELLDQARYFYSTGQSVRIHVQKNIW TYPLFVNTFSANALVG
35     Sbjct: 14802 TPVVACAVSKQSIWAPSFKELLDQARYFYSTGQSVRIHVQKNIW TYPLFVNTFSANALVG 14981

Query:      125 LSSCSATQCFGPK 137
              LSSCSATQCFGPK
40     Sbjct: 14982 LSSCSATQCFGPK 15020

      Score = 327 (120.2 bits), Expect = 1.1e-29, P = 1.1e-29
      Identities = 65/96 (67%), Positives = 73/96 (76%), Frame = +1

Query:      1 MYMSKYVPVY TLLILIYSFNASAEWTGDNTNAYYSDEVI SELHVGQIDTSPYFCIKTVKA 60
              MY++K+VPVY TLLILIYSFNASAEWTGDNTNAYYSDEVI SELHVGQIDTSPYFCIKTVKA
45     Sbjct: 14611 MYINKFVPVY TLLILIYSFNASAEWTGDNTNAYYSDEVI SELHVGQIDTSPYFCIKTVKA 14790

Query:      61 NGS GTPVVACAVSKQSIWAPSFKELLDQARYFYSTG 96
              NGS ++ ++ P K L + F G
50     Sbjct: 14791 NGS VHQLLHVRYQSRAYGRPLKNFLIRQDIFTVOG 14898

```

Toxin gene from *Streptococcus pyogenes*

A protein with the amino acid sequence <SEQ ID 7> was identified in *S. pyogenes*. This is encoded by a gene having nucleotide sequence <SEQ ID 8>.

55 This protein shows 24% identity to the C3 toxin from *Clostridium limosum*:

```

SEQID7      MLKKRYQLAIVLLLSCFSLIWQTEGLVELFVCEHYERAVCEGTP---AYFTFSQKGA
              | :| : | | | :| : : : | | : : | :| :|
exoc3_cloli. MNKLTERTVLCVGVSGLLIFSVAALVQGTKKCYANPVRNRAASRVKPYADSFKEFTNIDEA

5  SEQID7      ETLIKRWGKGLIYPRAEQEAMAAAYTCQQAGPINTSLDKAKGELSQLTPELRDQVAQLDA
              :: |::| : :|::|::| | :|::| | | :| : : | :| :| |::|
exoc3_cloli. RAWGDKQFAKYKL-SSSEKNALTIYT-RNAARINGPLRANQGNTNGLPADIRKEVEQIDK

10  SEQID7      ATHRLVIPWNIVVYRYVYETFLRDIGVSHADL--TSYYRNHQFDPHILCKIKL---GTRY
              : : : | |::| | | :| : | :| : : : :| | |
exoc3_cloli. SFTKMQTPENIILFRG-----DDPGYLGPDFENTILNRDGTINKAVFEQVKLRFKGKDR

15  SEQID7      TKHSFMSTTALKNGAMTHRPVEVRICVKKGAKAAAFVEPYSAVPSEVELLFPRGCQLEVVG
              ::::|::| :::|::| |::| : : | |::|::| | : : :|::| :
exoc3_cloli. KEYGYISTSLVNGSAFAGRPIITKFKVLDSKAGYIEPISTFKGQLEVLLPRSSTYTISD

20  SEQID7      AYVSQDQKKLHIEAYFKGSL
              : : :|::| | | :|
exoc3_cloli. MQIAPNNKQIIITALLKR

```

25 It also shows 29% identity to the EDIN transferase [M63917; ref. 23] of *S.aureus*:

Identities = 58/195 (29%), Positives = 106/195 (53%), Gaps = 13/195 (6%)

```

Query: 67  RWGKGLI----YPRAEQEAMAAAYTCQQAGPINTSLDKAKGELSQLTPELRDQVAQLDAAT 122
          +WG LI Y ++ A+ YT + + IN L A G++++L +D+V +LD++
30  Sbjet: 49  KWGNKLIKQAKYSSDDKIALYEYTKDSSKINGPLRLAGGDINKLDSTTQDKVRRDLSSI 107

```

```

Query: 123 HRLVIPWNIVVYRYVYETFLRDI-GVSHADLTSYYR--NHQFDPHILCKIK--LGTR-YT 176
          + P ++ VYR + +L I G ++ DL + N Q+D +++ K+ + +R Y
35  Sbjet: 108 SKSTTPESVYVYRLLNLDYLTISIVGFTNEDLYKLQQTNNQYDENLVRKLNNVMNSRIYR 167

```

```

Query: 177 KHSFMSTTALKNGAMTHRPVEVRICVKKGAKAAAFV--EPYSAVPSEVELLFPRGCQLEVV 234
          + + ST + A+ RP+E+R+ + KG KAA++ + +A + E+L PRG + V
40  Sbjet: 168 EDGYSSTQLVSGAAVGGRPIELRLLELPKGTKAAYLNSKDLTAYYGQQEVLLPRGTEYAVG 227

```

```

Query: 235 GAYVSQDQKKLHIEA 249
          +S D+KK+ I A
45  Sbjet: 228 SVELSNDKKKIIITA 242

```

Enzymatic studies

The *N.meningitidis* protein was expressed in and purified from *E.coli* as a His-tagged product. Mouse polyclonal antibodies raised against the recombinant protein were used in Western blot analysis and showed a sharp band at 20kDa in a cell lysate of *N. meningitidis* strain MC58. A preparation of outer membrane vesicles showed no such band.

As a preliminary assay for NAD-glycohydrolase activity, agmatine was used as an ADP-ribose acceptor. Purified protein was incubated in the presence of 0mM, 20mM or 75mM agmatine in 50mM potassium phosphate, 0.01mM [carbonyl-¹⁴C]NAD (0.05μCi), pH 7.5, in a total volume of 0.3 ml. After incubations at 30°C for 18hrs, samples (100μl) were applied to 1ml column of Dowex AG I-X2. [¹⁴C]Nicotinamide was eluted with 5 ml of H₂O for radioassay.

Enzymatic activity was as follows:

Agmatine concentration (mM)	Enzymatic activity (pmoles/hr)
0	34
20	62
75	72

Further studies of NAD-glycohydrolase and ADP-ribosyltransferase activity were performed using agmatine as acceptor. The NAD-glycohydrolase assay used 0.1M [carbonyl-¹⁴C]NAD, which was replaced with [adenine-U-¹⁴C]NAD for the ADP-ribosyltransferase assay. After incubation at 30°C for 1 hour, samples were tested as before. Results were as follows:

[Agmatine] (mM)	ADP-ribosylagmatine formed (nmol/hour)	Nicotinamide released (nmol/hour)
0	-	8.9
20	6.3	15.4
75	18.1	25.9

5 Further amino acids were tested as ADP-ribose acceptors at 20mM concentration. Results were:

Amino acid	Nicotinamide released (nmol/hour)
Control	8.2
Agmatine	15.6
Arginine	14.6
Glycine	8.5
Cysteine	11.7
Serine	8.6
Lysine	8.9
Histidine	8.8
Proline	8.6

Auto ADP-ribosylation of N.meningitidis toxin

Purified NMB1343 (5.7 µg) was incubated in 50 mM potassium phosphate (pH 7.5) with 10 µM [adenylate-³²P]NAD (10 µCi per assay) in a total volume of 50 µl for 1 h at 30°C. Protein was precipitated with the addition of 50 µl of ice-cold trichloroacetic acid (final concentration 25%) and, following an overnight incubation at 4°C, was collected by centrifugation (10,000 x g for 30 min). The pellet was suspended in LDS and heated at 70°C for 5 min. Samples were then subjected to electrophoresis in (4-12% or 10%) NuPAGE gels using MES as running buffer and electrotransferred to nitrocellulose membranes that were exposed to X-Omat film for 5 days at room temperature.

During incubation with [³²P]NAD, the radiolabel of NMB1343 increases in a concentration dependent manner (Figure 3A). To confirm that protein modification was enzymatic (*i.e.* that chemical addition of reactive ADP-ribose was not involved), labelling experiments were carried out

in the presence of Nam, a known NADase inhibitor and in the presence of a large excess of cold ADP-ribose. Both experiments gave an identical amount of radiolabelled protein (Figure 3B, L5 & L6). Enzymatic incorporation was also supported by the finding that pre-incubation of NMB1343 at 95°C for 5 minutes completely abolishes labelling (L1) and that labelling does not occur in the presence of novobiocin (L2), a known inhibitor of ADPRTs.

Mutant toxins

Based on homology with known toxins and the prediction of catalytic residues, four mutants SEQ ID 1 were generated, each containing a single amino acid substitution. The arginine in position 7 and the glutamic acids in positions 109, 111 and 120 were replaced with lysine and glycine residues respectively, using PCR-based site-directed mutagenesis (SDM). Internal primers containing a codon change from Arg to Lys, and from Glu to Gly were designed:

Primer	Sequence	{SEQ ID}	Codon change
WT-for	CGCGGATCCCATATGGGAAATTTCTTATATAGAGGCATTAGTTGC	{18}	
WT-rev	CCCGCTCGAGGTTAATTTCTATCAACTCTTTAGCAAT	{19}	
R7K-for	CGCGGATCCCATATGGGAAATTTCTTATAT <u>Aa</u> AGGCATTAGTTGC	{20}	AGA → AaA
E109G-for	ATTTTGAATAT <u>Gg</u> GGTTGAACATCCAGAAAAC	{21}	GAG → GgG
E109G-rev	TTCTGGATGTTCAACCCcCATATTCAAAAATAGA	{22}	
E111G-for	TATGAGGTT <u>Gg</u> ACATCCAGAAAACCCA	{23}	GAA → GgA
E111G-rev	GTTTTCTGGATGTcCAACCTCATATTC	{24}	
E120G-for	CCAAATGAGAAG <u>Gg</u> AGTAACAATCAGAG	{25}	GAA → GgA
E120G-rev	GATTGTTACTcCCTTCTCATTTGGGTT	{26}	

Underlined nucleotides code for lysine or glycine, and mutated nucleotides are in lower case.

To generate the R7K mutant, a single step of PCR was performed. The template was 20ng of DNA which encodes his-tagged NMB1343. Primers were R7K-for and WT-rev.

- 15 To generate the remaining mutants, PCR was performed using 20ng of pET 21b+ DNA as template, and the following primer pairs: (1) WT-for / E109G-rev; (2) E109G-for / WT-rev; (3) WT-for / E111G-rev; (4) E111G-for / WT-rev; (5) WT-for / E120G-rev; (6) E120G-for / WT-rev.

The second round of PCR was performed using the product of PCR 1-2, 3-4 or 5-6 as template, and WT-for and WT-rev as forward and reverse primers, respectively.

- 20 The PCR fragments containing each mutation were processed following standard procedure, digested with *Nde*I and *Xho*I restriction enzymes and cloned into pET-21b+ vector. The presence of each mutation was confirmed by sequence analysis.

- After cloning each gene into the expression vector, recombinant plasmids were transformed into *E.coli* strains suitable for expression of the recombinant protein as His-tag. 1.5µl of each construct was used to transform *E.coli* BL21-DE3. Single recombinant colonies were inoculated into 4ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp

(100 μ g/ml) in 125ml flasks, to give an OD₆₀₀ between 0.1 and 0.2. The flasks were incubated at 37°C in a gyratory water bath shaker until OD₆₀₀ indicated exponential growth suitable for induction of expression (0.4-0.8 OD). Protein expression was induced by addition of 1.0mM IPTG. After 3 hours incubation at 37°C the OD₆₀₀ was measured and expression examined. 1.0ml of each sample was centrifuged in a microfuge, the pellet resuspended in PBS and analysed by SDS-PAGE and Coomassie Blue staining. All the mutants were expressed as efficiently as wild-type, and they were purified as soluble forms.

ADP-ribosyltransferase and NAD-glycohydrolase activities were determined as described above. The Lys and Gly mutations were studied separately. Results were as follows:

Mutant	ADP-ribosyltransferase activity (pmol/hour)	NAD-glycohydrolase activity (nmol/hour)
Wild type	1.8	2.6
R7K	0.1	0.08
Wild type	305	680
E109G	30	315
E111G	116	289
E120G	36	100

Thus the enzymatic activity of the *N.meningitidis* toxin can be rationally and efficiently removed by site-directed mutagenesis.

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

TABLE 1 – preferred mutations (Δ = deletion of the residue)

SEQ ID	Site for mutation(s)	Replacement residue
1	Arg-7	Lys, Δ
	Lys-24	Gly, Trp
	Tyr-34	Trp, Ala, His
	His-57	Asn, Tyr, Gln, Val, Ser, Pro
	Glu-60	Ala
	Thr-61	Gly
	Tyr-68	Met, Glu
	Ser-70	Phe
	Thr-72	Lys, Tyr
	Ala-82	Arg
	Gly-86	Lys
	Tyr-103	Lys, Asp, Ser
	Glu-107	Asp, Ser, Δ
	Glu-109	Ala, Gly, Lys, Asp, Ser, Δ
	Glu-111	Ala, Gly, Lys, Asp, Gln, Δ
3	Glu-118	Asp, Ser, Δ
	Glu-120	Ala, Gly, Lys, Asp, Gln, Δ
	Asp-64	Glu, Tyr
	Arg-81	Lys, Gly, Trp, Δ
	Asp-83	Glu, Tyr, Ser
	Arg-85	Lys, His, Leu
	Gly-86	Arg, Asp
	His-96	Asn, Tyr, Gln, Val, Ser, Pro
	Gln-101	Ala
	Tyr-110	Met
	Val-113	Asp, Glu, Tyr
	Tyr-119	Glu
	Ser-121	Phe
	Ser-123	Lys, Tyr
	His-126	Asn, Tyr, Gln, Val
4	Gly-136	Lys
	Val-149	Lys, Tyr
	His-157	Glu
	Gln-162	Asp, Ser, Δ
	Glu-164	Ala, Gly, Lys, Asp, Gln, Δ
	Arg-10	Lys, Δ
	Asp-12	Glu, Tyr, Ser
	Arg-14	Lys, His, Leu
	Ile-19	Ala
	His-36	Asn, Tyr, Gln, Val, Ser, Pro
	Thr-40	Gly
	Arg-44	Ala, Lys

	Phe-47 Ser-49 Ser-51 Ala-57 Gln-130 Glu-132 Trp-133	Met, Glu Phe Lys, Tyr Arg Asp, Ser, Δ Ala, Gly, Lys, Asp, Gln, Δ Gly
5/6	Arg-24 Asp-26 Ile-33 Tyr-43 Arg-46 Cys-56 Ser-61 Ser-63 Tyr-65 Thr-68 Thr-69 Ser-70 Ser-84 Ser-86 Gln-131 Glu-133 Tyr-134	Lys, Δ Glu, Tyr, Ser Ala Trp, Ala, His Ala, Lys Δ Phe Lys, Tyr Met, Glu Phe Phe Lys, Tyr Phe Lys, Tyr Asp, Ser, Δ Gly, Ala, Lys, Asp, Gln, Δ Gly
7	Arg-130 Ser-177 Thr-179 Cys-196 Glu-215 Glu-217	Lys, Δ Phe Lys, Tyr Ser Ser, Asp, Δ Gly, Ala, Lys, Asp, Gln, Δ

REFERENCES (the contents of which are hereby incorporated in full)

- [1] Rappuoli & Pizza (1991) Chapter 1 of *Sourcebook of Bacterial Protein Toxins* (Alouf & Freer, eds). ISBN 0-12-053078-3.
- [2] Bazan & Koch-Nolte (1997) *Adv. Exp. Med. Biol.* 419:99-107.
- [3] Sixma *et al.* (1991) *Nature* 351:371-377.
- [4] Zhang *et al.* (1995) *J. Mol. Biol.* 251:563-573.
- [5] Stein *et al.* (1994) *Structure* 2:45-57.
- [6] International patent application WO93/13202.
- [7] European patent applications 0306618, 0322533 and 0322115.
- [8] Del Giudice & Rappuoli (1999) *Vaccine* 1999 17 Suppl 2:S44-52
- [9] European patent 0396964.
- [10] Northrup & Fauci (1972) *J. Infect. Dis.* 125:672ff.
- [11] Elson & Ealading (1984) *J. Immunol.* 133:2892ff and 132:2736ff.
- [12] International patent application WO95/17211.
- [13] van den Akker *et al.* (1997) *Protein Sci* 6:2644-2649.
- [14] International patent application WO99/24578.
- [15] International patent application WO99/36544.
- [16] International patent application WO99/57280.
- [17] Tettelin *et al.* (2000) *Science* 287:1809-1815.
- [18] Pizza *et al.* (2000) *Science* 287:1816-1820.
- [19] International patent application WO01/64920.
- [20] International patent application WO01/64922.
- [21] Parkhill *et al.* (2000) *Nature* 404:502-506.
- [22] Himmelreich *et al.* (1996) *Nucleic Acids Res.* 24:4420-4449.
- [23] Wilde *et al.* (2001) *J. Biol. Chem.* 276:9537-9542.

CLAIMS

1. A protein consisting of an amino acid sequence of one of SEQ IDs 1, 3, 4, 5, 6 or 7.
2. A protein consisting of an amino acid sequence of one of SEQ IDs 1, 3, 4, 5, 6 or 7, except that the amino acid sequence contains one or more mutations.
- 5 3. The protein of claim 2, wherein the amino acid sequence contains between one and twenty mutations.
4. The protein of claim 2 or claim 3, wherein the mutation(s) reduce or eliminate ADP-ribosylating activity of the protein.
5. The protein of any one of claims 2 to 4, wherein the mutation(s) are independently
10 substitutions, insertions, and/or deletions.
6. The protein of claim 5, wherein the mutation(s) include single amino acid substitutions.
7. The protein of claim 5 or claim 6, wherein the mutation comprises one or more of the mutations listed in Table 1.
8. The protein of claim 7, comprising SEQ ID 10, 12, 14 and/or 16.
- 15 9. A protein comprising the amino acid sequence of a protein according to any preceding claim.
10. A protein comprising an amino acid sequence having greater than 50% sequence identity to the amino acid sequence of a protein according to any one of claims 1 to 8.
11. A protein comprising a fragment of 7 or more amino acids from a protein according to any one of claims 1 to 8.
- 20 12. The protein of any preceding claim, for use as an adjuvant.
13. The protein of claim 12, for use as a mucosal adjuvant.
14. The use of the protein of any one of claims 1 to 11 in the manufacture of a medicament for raising an immune response in an animal.
15. An immunogenic composition comprising the protein of any one of claims 1 to 11 in
25 admixture with an antigen.
16. Antibody which binds to the protein of any one of claims 1 to 11.
17. Nucleic acid encoding the protein of any one of claims 1 to 11.

18. Nucleic acid comprising a fragment of 10 or more nucleotides from the nucleic acid of claim 17.
19. A vector comprising a nucleic acid sequence of the nucleic acid of claim 17 or claim 18.
20. A host cell transformed with the vector of claim 19.
- 5 21. A composition comprising the protein of any one of claims 1 to 11, the antibody of claim 16, and/or the nucleic acid of claim 17 or claim 18.
22. The composition of claim 21 for use as a medicament.
23. Use of the composition of claim 21 in the manufacture of a medicament for treating or preventing bacterial infection.
- 10 24. The use of claim 21, wherein the bacterial infection is caused by *Neisseria meningitidis*, *Streptomyces coelicolor*, *Mycoplasma pneumoniae*, *Salmonella typhimurium*, *Salmonella paratyphi*, or *Streptococcus pyogenes*.
25. A method of treating a patient, comprising administering to the patient a therapeutically effective amount of the composition of claim 21.
- 15 26. Nucleic acid comprising one or more of the nucleotide sequences SEQ IDs 2, 8, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25 and/or 26.

FIGURE 1



CT group

Region 1



CT	IYRA ⁸
NMB1343	IYRA ⁸
LT	IYRA ⁸
LT-II	FFRA ⁶
PT	VYRY ¹⁰
Exos	...
C2_bot	AYRR ¹⁰⁰
C3_bot	LFRG ⁸⁹
C3-II	IYRG ⁸⁹
C3_lim	LFRG ⁸⁹
IOTA_tox	VYRR ²⁹⁷
B. cereus	...
Edin	VYRL ¹²¹

Region 2



55	HDDG	YVST	ISL	RS	AHL	VG	QTI	⁷⁶
64	YDGC	YIST	TT	TD	KE	IA	KK	⁸²
55	YDDG	YVST	SL	SL	RS	AHL	AG	⁷⁶
53	YNDG	YVST	TV	TL	RQ	AHL	IG	⁷⁴
53	FVST	SS	SR	RY	TE	VY	LE	⁷⁶
338	DDGY	LS	SL	NP	GV	AR	SG	³⁵⁸
345	SFST	SL	KL	ST	PL	SF	S	³⁵⁹
131	GYIS	ST	SL	MS	AQ	FG	GR	¹⁴⁵
131	GYIS	ST	SL	MN	VS	QF	AG	¹⁴⁵
131	GYIS	ST	SL	NG	SA	FA	GR	¹⁴⁷
336	FIST	SI	GS	VN	MS	AF	AK	³⁹⁴
170	GYSS	ST	QL	VS	GA	AV	GG	¹⁸⁵

Region 3



107	HPD	EQ	EV	SAL	¹¹⁶
107	HPY	EQ	EV	SAL	¹¹⁶
106	YPS	EN	EA	AL	¹¹⁵
126	TYQ	SE	YL		¹³⁰
376	YKN	KE	IL	LY	³⁸⁴
385	EQ	EL	LN		³⁹³
169	FPG	QL	EV	LPL	¹⁷⁸
169	FAG	QL	EM	LPL	¹⁷⁸
169	FKG	QL	EV	LPL	¹⁷⁸
375	YAG	YE	VL	L	³⁸³
...	YPG	QY	EL	LPL	...
210	YYG	QY	EL	LPL	²¹⁹



DT group

DT	SYHBT ²³
PAETA	GYHGT ⁴⁴²

50	WRGFY	STD	NK	YD	AA	SY	⁵⁵
466	WRGFY	IAG	DP	AL	AY	SY	¹⁸¹

14	EYIN
55	ETIL

FIGURE 2

<i>E. coli</i> LT	KLYRADSRPPD	LYDHARGTQ	YDDGYVSTSLRSA	SPHPYEQEVSA LGGI
<i>N. meningitidis</i>	ELYRGISCQQD	VYAHQIETG	YDGCYISTTTDKELA	PENPNEKEVITRAED
<i>S. coelicolor</i>	ELYRSDSRGPQ	VVFEEGFHAKDVQNGQYDVEKYVLVNQSPVYSTSYDHDLY		HKWADQVEVAFPPGI
<i>M. pneumoniae</i>	FVYRVDSRPE	FFEHILSTN	GRSYFISTSEPTAA	TSFAYQRENTDGP I
<i>S. typhi</i>	FVYRVDS TPPD		SCSGGSDSRVYIATT	TMMRLQREYVSTLSI
<i>S. paratyphi</i>	FVYRVDS TPPD		SCSGGSDSRVYIATT	TMMRLQREYVSTLSI
<i>S. pyogenes</i>	VYRVVYETFL		TKHSFMSTTALKNGA	SAVPSEVELEPRGC

FIGURE 3B

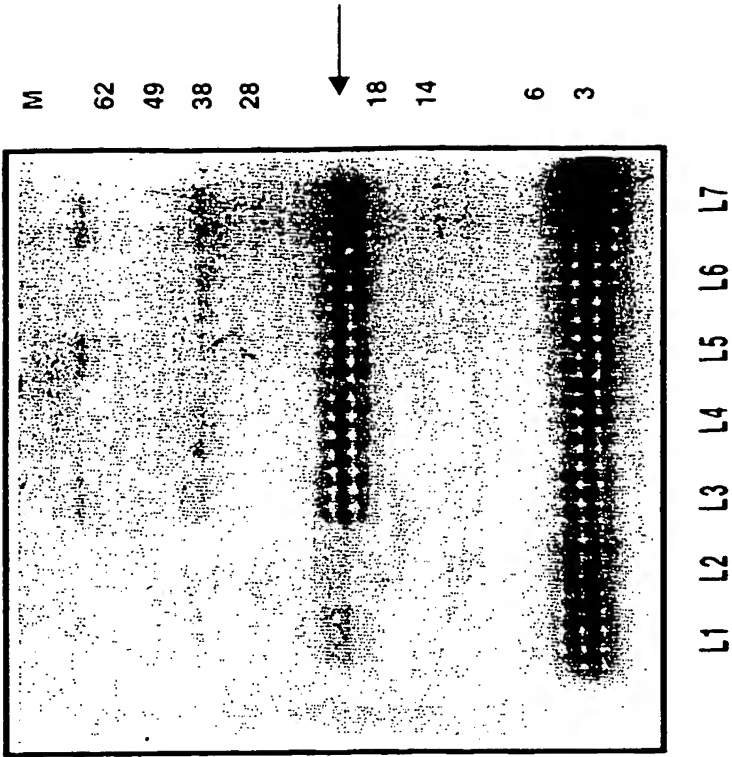
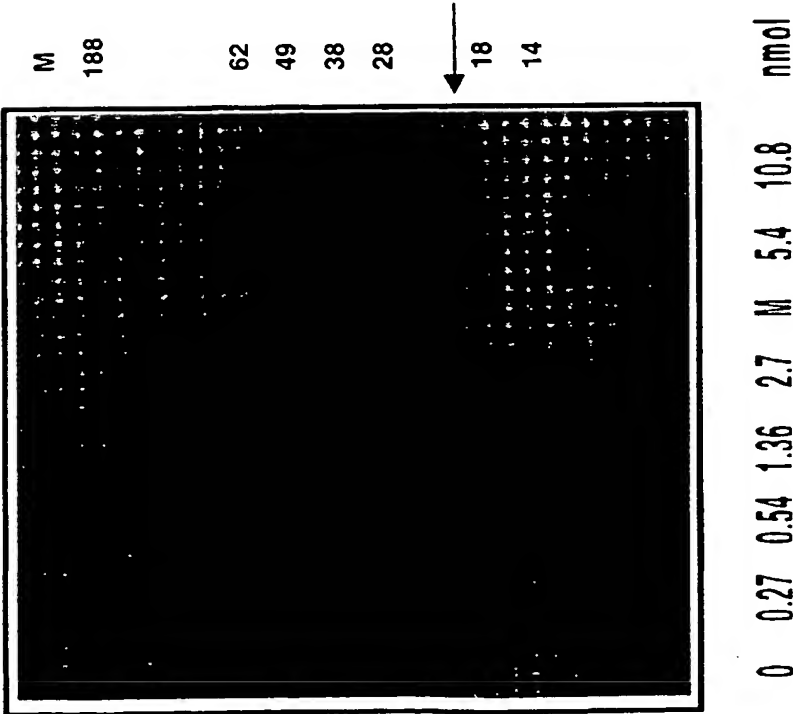


FIGURE 3A



SEQUENCE LISTING

<110> CHIRON SpA

<120> ADP-RIBOSYLATING BACTERIAL TOXINS

<130> P026968WO

<150> GB-0108024.1

<151> 2001-03-30

<160> 26

<170> SeqWin99, version 1.02

<210> 1

<211> 145

<212> PRT

<213> Neisseria meningitidis

<400> 1

Met Gly Asn Phe Leu Tyr Arg Gly Ile Ser Cys Gln Gln Asp Glu Gln

1 5 10 15

Asn Asn Gly Gln Leu Lys Pro Lys Gly Asn Lys Ala Glu Val Ala Ile

20 25 30

Arg Tyr Asp Gly Lys Phe Lys Tyr Asp Gly Lys Ala Thr His Gly Pro

35 40 45

Ser Val Lys Asn Ala Val Tyr Ala His Gln Ile Glu Thr Gly Leu Tyr

50 55 60

Asp Gly Cys Tyr Ile Ser Thr Thr Thr Asp Lys Glu Ile Ala Lys Lys

65 70 75 80

Phe Ala Thr Ser Ser Gly Ile Glu Asn Gly Tyr Ile Tyr Val Leu Asn

85 90 95

Arg Asp Leu Phe Gly Gln Tyr Ser Ile Phe Glu Tyr Glu Val Glu His

100 105 110

Pro Glu Asn Pro Asn Glu Lys Glu Val Thr Ile Arg Ala Glu Asp Cys

115 120 125

Gly Cys Ile Pro Glu Glu Val Ile Ile Ala Lys Glu Leu Ile Glu Ile

130 135 140

Asn

145

<210> 2

<211> 435

<212> DNA

<213> Neisseria meningitidis

<400> 2

atgggaaatt tcttatatag aggcattagt tgccaacaag atgagcaaaa taatggacag 60

ttaaaaccta aaggtaataa agctgaagtt gcaattcggt atgatggtaa gtttaaatat 120

gatggtaaag ctacacatgg tccaagtgtg aagaatgcag ttacgcca tcaaattgaa 180

acaggtctat atgacggatg ttatatatct acgacaacag acaaggaaat tgccaagaaa 240

-2-

tttgcaacaa gttccggcat cgaaaatggc tatatatatg ttttaaataag ggatttggtt 300
 ggtcaatatt ctatttttga atatgaggtt gaacatccag aaaacccaaa tgagaaggaa 360
 gtaacaatca gagctgaaga ttgtggctgt attcctgaag aagtgattat tgctaaagag 420
 ttgatagaaa ttaac 435

5
 <210> 3
 <211> 204
 <212> PRT
 <213> Streptomyces coelicolor

10
 <400> 3
 Met Ile Thr Thr Ser Leu Arg Arg Arg Thr Ala Ala Ala Val Leu Ser
 1 5 10 15

15 Leu Ser Ala Val Leu Ala Thr Thr Ala Ala Thr Ala Pro Gly Ala Ala
 20 25 30

Pro Ala Pro Ser Ala Ala Pro Ala Lys Ala Ala Pro Ala Cys Pro Gln
 35 40 45

20 Phe Asp Asp Arg Thr Lys Ala Ala Ala Asp Arg Gly Val Asp Val Asp
 50 55 60

25 Arg Ile Thr Pro Glu Pro Val Trp Arg Thr Thr Cys Gly Thr Leu Tyr
 65 70 75 80

Arg Ser Asp Ser Arg Gly Pro Gln Val Val Phe Glu Glu Gly Phe His
 85 90 95

30 Ala Lys Asp Val Gln Asn Gly Gln Tyr Asp Val Glu Lys Tyr Val Leu
 100 105 110

Val Asn Gln Pro Ser Pro Tyr Val Ser Thr Ser Tyr Asp His Asp Leu
 115 120 125

35 Tyr Lys Thr Trp Tyr Lys Ser Gly Tyr Asn Tyr Tyr Val Asp Ala Pro
 130 135 140

40 Gly Gly Ile Asp Val Asn Lys Thr Ile Gly Asp Thr His Lys Trp Ala
 145 150 155 160

Asp Gln Val Glu Val Ala Phe Pro Gly Gly Ile Gln Arg Lys Tyr Ile
 165 170 175

45 Ile Gly Val Cys Pro Val Asp Arg Gln Thr Lys Thr Glu Ile Met Ser
 180 185 190

Asp Cys Glu Ser Asn Pro His Tyr Gln Pro Trp His
 195 200

50
 <210> 4
 <211> 591
 <212> PRT
 <213> Mycoplasma pneumoniae

55
 <400> 4
 Met Pro Asn Pro Val Arg Phe Val Tyr Arg Val Asp Leu Arg Ser Pro
 1 5 10 15

60 Glu Glu Ile Phe Glu His Gly Phe Ser Thr Leu Gly Asp Val Arg Asn
 20 25 30

Phe Phe Glu His Ile Leu Ser Thr Asn Phe Gly Arg Ser Tyr Phe Ile
 .. 35 40 45
 5 Ser Thr Ser Glu Thr Pro Thr Ala Ala Ile Arg Phe Phe Gly Ser Trp
 50 55 60
 Leu Arg Glu Tyr Val Pro Glu His Pro Arg Arg Ala Tyr Leu Tyr Glu
 65 70 75 80
 10 Ile Arg Ala Asp Gln His Phe Tyr Asn Ala Arg Ala Thr Gly Glu Asn
 85 90 95
 Leu Leu Asp Leu Met Arg Gln Arg Gln Val Val Phe Asp Ser Gly Asp
 15 100 105 110
 Arg Glu Met Ala Gln Met Gly Ile Arg Ala Leu Arg Thr Ser Phe Ala
 115 120 125
 20 Tyr Gln Arg Glu Trp Phe Thr Asp Gly Pro Ile Ala Ala Ala Asn Val
 130 135 140
 Arg Ser Ala Trp Leu Val Asp Ala Val Pro Val Glu Pro Gly His Ala
 145 150 155 160
 25 His His Pro Ala Gly Arg Val Val Glu Thr Thr Arg Ile Asn Glu Pro
 165 170 175
 Glu Met His Asn Pro His Tyr Gln Glu Leu Gln Thr Gln Ala Asn Asp
 30 180 185 190
 Gln Pro Trp Leu Pro Thr Pro Gly Ile Ala Thr Pro Val His Leu Ser
 195 200 205
 35 Ile Pro Gln Ala Ala Ser Val Ala Asp Val Ser Glu Gly Thr Ser Ala
 210 215 220
 Ser Leu Ser Phe Ala Cys Pro Asp Trp Ser Pro Pro Ser Ser Asn Gly
 225 230 235 240
 40 Glu Asn Pro Leu Asp Lys Cys Ile Ala Glu Lys Ile Asp Asn Tyr Asn
 245 250 255
 Leu Gln Ser Leu Pro Gln Tyr Ala Ser Ser Val Lys Glu Leu Glu Asp
 45 260 265 270
 Thr Pro Val Tyr Leu Arg Gly Ile Lys Thr Gln Lys Thr Phe Met Leu
 275 280 285
 50 Gln Ala Asp Pro Gln Asn Asn Asn Val Phe Leu Val Glu Val Asn Pro
 290 295 300
 Lys Gln Lys Ser Ser Phe Pro Gln Thr Ile Phe Phe Trp Asp Val Tyr
 305 310 315 320
 55 Gln Arg Ile Cys Leu Lys Asp Leu Thr Gly Ala Gln Ile Ser Leu Ser
 325 330 335
 Leu Thr Ala Phe Thr Thr Gln Tyr Ala Gly Gln Leu Lys Val His Leu
 60 340 345 350

-4-

Ser Val Ser Ala Val Asn Ala Val Asn Gln Lys Trp Lys Met Thr Pro
 355 360 365
 5 Gln Asp Ile Ala Ile Thr Gln Phe Arg Val Ser Ser Glu Leu Leu Gly
 370 375 380
 Gln Thr Glu Asn Gly Leu Phe Trp Asn Thr Lys Ser Gly Gly Ser Gln
 385 390 395 400
 10 His Asp Leu Tyr Val Cys Pro Leu Lys Asn Pro Pro Ser Asp Leu Glu
 405 410 415
 Glu Leu Gln Ile Ile Val Asp Glu Cys Thr Thr His Ala Gln Phe Val
 420 425 430
 15 Thr Met Arg Ala Ala Ser Thr Phe Phe Val Asp Val Gln Leu Gly Trp
 435 440 445
 Tyr Trp Arg Gly Tyr Tyr Tyr Thr Pro Gln Leu Ser Gly Trp Ser Tyr
 450 455 460
 Gln Met Lys Thr Pro Asp Gly Gln Ile Phe Tyr Asp Leu Lys Thr Ser
 465 470 475 480
 25 Lys Ile Phe Phe Val Gln Asp Asn Gln Asn Val Phe Phe Leu His Asn
 485 490 495
 Lys Leu Asn Lys Gln Thr Gly Tyr Ser Trp Asp Trp Val Glu Trp Leu
 500 505 510
 30 Lys His Asp Met Asn Glu Asp Lys Asp Glu Asn Phe Lys Trp Tyr Phe
 515 520 525
 Ser Arg Asp Asp Leu Thr Ile Pro Ser Val Glu Gly Leu Asn Phe Arg
 530 535 540
 His Ile Arg Cys Tyr Ala Asp Asn Gln Gln Leu Lys Val Ile Ile Ser
 545 550 555 560
 40 Gly Ser Arg Trp Gly Gly Trp Tyr Ser Thr Tyr Asp Lys Val Glu Ser
 565 570 575
 Asn Val Glu Asp Lys Ile Leu Val Lys Asp Gly Phe Asp Arg Phe
 580 585 590
 45 <210> 5
 <211> 242
 <212> PRT
 <213> Salmonella typhimurium
 50 <400> 5
 Met Lys Lys Leu Ile Phe Leu Thr Leu Ser Ile Val Ser Phe Asn Asn
 1 5 10 15
 55 Tyr Ala Val Asp Phe Val Tyr Arg Val Asp Ser Thr Pro Pro Asp Val
 20 25 30
 Ile Phe Arg Asp Gly Phe Ser Leu Leu Gly Tyr Asn Arg Asn Phe Gln
 35 40 45
 60 Gln Phe Ile Ser Gly Arg Ser Cys Ser Gly Gly Ser Ser Asp Ser Arg

-5-

50 55 60
 Tyr Ile Ala Thr Thr Ser Ser Val Asn Gln Thr Tyr Ala Ile Ala Arg
 65 70 75 80
 5 Ala Tyr Tyr Ser Arg Ser Thr Phe Lys Gly Asn Leu Tyr Arg Tyr Gln
 85 90 95
 10 Ile Arg Ala Asp Asn Asn Phe Tyr Ser Leu Leu Pro Ser Ile Thr Tyr
 100 105 110
 Leu Glu Thr Gln Gly Gly His Phe Asn Ala Tyr Glu Lys Thr Met Met
 115 120 125
 15 Arg Leu Gln Arg Glu Tyr Val Ser Thr Leu Ser Ile Leu Pro Glu Asn
 130 135 140
 20 Ile Gln Lys Ala Val Ala Leu Val Tyr Asp Ser Ala Thr Gly Leu Val
 145 150 155 160
 Lys Asp Gly Val Ser Thr Met Asn Ala Ser Tyr Leu Gly Leu Ser Thr
 165 170 175
 25 Thr Ser Asn Pro Gly Val Ile Pro Phe Leu Pro Glu Pro Gln Thr Tyr
 180 185 190
 Thr Gln Gln Arg Ile Xaa Ala Phe Gly Pro Leu Ile Ser Ser Cys Phe
 195 200 205
 30 Ser Ile Gly Ser Val Cys His Ser His Arg Gly Gln Arg Ala Asp Val
 210 215 220
 Tyr Asn Met Ser Phe Tyr Asp Ala Arg Pro Val Ile Glu Leu Ile Leu
 225 230 235 240
 35 Ser Lys
 40 <210> 6
 <211> 242
 <212> PRT
 <213> Salmonella paratyphi
 45 <400> 6
 Met Lys Lys Leu Ile Phe Leu Thr Leu Ser Ile Val Ser Phe Asn Asn
 1 5 10 15
 Tyr Ala Val Asp Phe Val Tyr Arg Val Asp Ser Thr Pro Pro Asp Val
 20 25 30
 50 Ile Phe Arg Asp Gly Phe Ser Leu Leu Gly Tyr Asn Arg Asn Phe Gln
 35 40 45
 Gln Phe Ile Ser Gly Arg Ser Cys Ser Gly Gly Ser Ser Asp Ser Arg
 55 50 55 60
 Tyr Ile Ala Thr Thr Ser Ser Val Asn Gln Thr Tyr Ala Ile Ala Arg
 65 70 75 80
 60 Ala Tyr Tyr Ser Arg Ser Thr Phe Lys Gly Asn Leu Tyr Arg Tyr Gln
 85 90 95

-6-

Ile Arg Ala Asp Asn Asn Phe Tyr Ser Leu Leu Pro Ser Ile Thr Tyr
 100 105 110
 5 Leu Glu Thr Gln Gly Gly His Phe Asn Ala Tyr Glu Lys Thr Met Met
 115 120 125
 Arg Leu Gln Arg Glu Tyr Val Ser Thr Leu Ser Ile Leu Pro Glu Asn
 130 135 140
 10 Ile Gln Lys Ala Val Ala Leu Val Tyr Asp Ser Ala Thr Gly Leu Val
 145 150 155 160
 Lys Asp Gly Val Ser Thr Met Asn Ala Ser Tyr Leu Gly Leu Ser Thr
 165 170 175
 15 Thr Ser Asn Pro Gly Val Ile Pro Phe Leu Pro Glu Pro Gln Thr Tyr
 180 185 190
 20 Thr Gln Gln Arg Ile Asp Ala Phe Gly Pro Leu Ile Ser Ser Cys Phe
 195 200 205
 Ser Ile Gly Ser Val Cys Gln Ser His Arg Gly Gln Arg Ala Asp Val
 210 215 220
 25 Tyr Asn Met Ser Phe Tyr Asp Ala Arg Pro Val Ile Glu Leu Ile Leu
 225 230 235 240
 Ser Lys
 30
 <210> 7
 <211> 250
 <212> PRT
 35 <213> Streptococcus pyogenes
 <400> 7
 Met Leu Lys Lys Arg Tyr Gln Leu Ala Ile Val Leu Leu Leu Ser Cys
 1 5 10 15
 40 Phe Ser Leu Ile Trp Gln Thr Glu Gly Leu Val Glu Leu Phe Val Cys
 20 25 30
 Glu His Tyr Glu Arg Ala Val Cys Glu Gly Thr Pro Ala Tyr Phe Thr
 35 40 45
 Phe Ser Asp Gln Lys Gly Ala Glu Thr Leu Ile Lys Lys Arg Trp Gly
 50 55 60
 50 Lys Gly Leu Ile Tyr Pro Arg Ala Glu Gln Glu Ala Met Ala Ala Tyr
 65 70 75 80
 Thr Cys Gln Gln Ala Gly Pro Ile Asn Thr Ser Leu Asp Lys Ala Lys
 85 90 95
 55 Gly Glu Leu Ser Gln Leu Thr Pro Glu Leu Arg Asp Gln Val Ala Gln
 100 105 110
 Leu Asp Ala Ala Thr His Arg Leu Val Ile Pro Trp Asn Ile Val Val
 115 120 125
 60

-7-

Tyr Arg Tyr Val Tyr Glu Thr Phe Leu Arg Asp Ile Gly Val Ser His
 130 135 140

5 Ala Asp Leu Thr Ser Tyr Tyr Arg Asn His Gln Phe Asp Pro His Ile
 145 150 155 160

Leu Cys Lys Ile Lys Leu Gly Thr Arg Tyr Thr Lys His Ser Phe Met
 165 170 175

10 Ser Thr Thr Ala Leu Lys Asn Gly Ala Met Thr His Arg Pro Val Glu
 180 185 190

Val Arg Ile Cys Val Lys Lys Gly Ala Lys Ala Ala Phe Val Glu Pro
 195 200 205

15 Tyr Ser Ala Val Pro Ser Glu Val Glu Leu Leu Phe Pro Arg Gly Cys
 210 215 220

20 Gln Leu Glu Val Val Gly Ala Tyr Val Ser Gln Asp Gln Lys Lys Leu
 225 230 235 240

His Ile Glu Ala Tyr Phe Lys Gly Ser Leu
 245 250

25 <210> 8
 <211> 765
 <212> DNA
 <213> Streptococcus pyogenes

30 <400> 8
 gtgtcagggg gaactatgct aaaaaagcgc tatcaactgg ctattgtcct tcttcttagc 60
 tgttttagtc tgatctggca aactgagggc ttggtcgagc tttttgtctg tgagcactat 120
 gagcgggcgg tttgtgaggg gacgcctgct tattttacct ttccggatca aaagggcgct 180
 gagacactga ttaaaaaagcg atggggcaag ggtctcatct acccaagggc tgagcaagag 240
 35 gcgatggctg cttataacct tcagcaggca ggcctatca acaccagcct agacaaagcc 300
 aaagggtgagc tcagccaact cacgcctgag ctaagggatc aggtggccca gctcgatgct 360
 gcgactcacc ggctagtcac cccgtggaac attgtagtat accgctatgt atacgagacg 420
 tttttgcgtg atatcggtgt ttcacatgct gatctcacgt cttactaccg taaccatcag 480
 tttgaccctc atatcctttg taagatcaag cttggtacac gctacaccaa gcacagtttt 540
 40 atgagcacga cagccttgaa aaacggcgcc atgaccatc gaccgggtgga ggtgcgcac 600
 tgtgtcaaaa aagggggccaa ggcagccttt gtcgagcctt attcggtgtg gccttcagag 660
 gttgagctct tgtttccaag aggtgtgcag ctggaggtcg ttggagctta cgtgtcacag 720
 gaccaaaaaa agctccacat agaagcgtat ttcaagggca gttttg 765

45 <210> 9
 <211> 137
 <212> PRT
 <213> Salmonella typhimurium

50 <400> 9
 Met Tyr Met Ser Lys Tyr Val Pro Val Tyr Thr Leu Leu Ile Leu Ile
 1 5 10 15

55 Tyr Ser Phe Asn Ala Ser Ala Glu Trp Thr Gly Asp Asn Thr Asn Ala
 20 25 30

Tyr Tyr Ser Asp Glu Val Ile Ser Glu Leu His Val Gly Gln Ile Asp
 35 40 45

60 Thr Ser Pro Tyr Phe Cys Ile Lys Thr Val Lys Ala Asn Gly Ser Gly
 50 55 60

-8-

Thr Pro Val Val Ala Cys Ala Val Ser Lys Gln Ser Ile Trp Ala Pro
 65 70 75 80
 5 Ser Phe Lys Glu Leu Leu Asp Gln Ala Arg Tyr Phe Tyr Ser Thr Gly
 85 90 95
 Gln Ser Val Arg Ile His Val Gln Lys Asn Ile Trp Thr Tyr Pro Leu
 100 105 110
 10 Phe Val Asn Thr Phe Ser Ala Asn Ala Leu Val Gly Leu Ser Ser Cys
 115 120 125
 Ser Ala Thr Gln Cys Phe Gly Pro Lys
 15 130 135
 <210> 10
 <211> 145
 <212> PRT
 20 <213> Neisseria meningitidis
 <400> 10
 Met Gly Asn Phe Leu Tyr Lys Gly Ile Ser Cys Gln Gln Asp Glu Gln
 1 5 10 15
 25 Asn Asn Gly Gln Leu Lys Pro Lys Gly Asn Lys Ala Glu Val Ala Ile
 20 25 30
 Arg Tyr Asp Gly Lys Phe Lys Tyr Asp Gly Lys Ala Thr His Gly Pro
 30 35 40 45
 Ser Val Lys Asn Ala Val Tyr Ala His Gln Ile Glu Thr Gly Leu Tyr
 50 55 60
 35 Asp Gly Cys Tyr Ile Ser Thr Thr Thr Asp Lys Glu Ile Ala Lys Lys
 65 70 75 80
 Phe Ala Thr Ser Ser Gly Ile Glu Asn Gly Tyr Ile Tyr Val Leu Asn
 85 90 95
 40 Arg Asp Leu Phe Gly Gln Tyr Ser Ile Phe Glu Tyr Glu Val Glu His
 100 105 110
 Pro Glu Asn Pro Asn Glu Lys Glu Val Thr Ile Arg Ala Glu Asp Cys
 45 115 120 125
 Gly Cys Ile Pro Glu Glu Val Ile Ile Ala Lys Glu Leu Ile Glu Ile
 130 135 140
 50 Asn
 145
 <210> 11
 <211> 435
 55 <212> DNA
 <213> Neisseria meningitidis
 <400> 11
 atgggaaatt tcttatataa aggcattagt tgccaacaag atgagcaaaa taatggacag 60
 60 ttaaaaccta aaggtaataa agctgaagtt gcaattcgtt atgatggtaa gtttaaatat 120
 gatggtaaag ctacacatgg tccaagtgtg aagaatgcag tttacgcca tcaaattgaa 180

-9-

acaggtctat atgacggatg ttatatatct acgacaacag acaaggaaat tgccaagaaa 240
 ttgcaacaa gttccggcat cgaaaatggc tatatatatg ttttaaatag ggatttggtt 300
 ggtcaatatt ctatttttga atatgaggtt gaacatccag aaaacccaaa tgagaaggaa 360
 gtaacaatca gagctgaaga ttgtggctgt attcctgaag aagtgattat tgctaaagag 420
 5 ttgatagaaa ttaac 435

<210> 12
 <211> 145
 <212> PRT
 10 <213> Neisseria meningitidis

<400> 12
 Met Gly Asn Phe Leu Tyr Arg Gly Ile Ser Cys Gln Gln Asp Glu Gln
 1 5 10 15
 15 Asn Asn Gly Gln Leu Lys Pro Lys Gly Asn Lys Ala Glu Val Ala Ile
 20 25 30
 20 Arg Tyr Asp Gly Lys Phe Lys Tyr Asp Gly Lys Ala Thr His Gly Pro
 35 40 45
 Ser Val Lys Asn Ala Val Tyr Ala His Gln Ile Glu Thr Gly Leu Tyr
 50 55 60
 25 Asp Gly Cys Tyr Ile Ser Thr Thr Thr Asp Lys Glu Ile Ala Lys Lys
 65 70 75 80
 Phe Ala Thr Ser Ser Gly Ile Glu Asn Gly Tyr Ile Tyr Val Leu Asn
 85 90 95
 30 Arg Asp Leu Phe Gly Gln Tyr Ser Ile Phe Glu Tyr Gly Val Glu His
 100 105 110
 Pro Glu Asn Pro Asn Glu Lys Glu Val Thr Ile Arg Ala Glu Asp Cys
 115 120 125
 35 Gly Cys Ile Pro Glu Glu Val Ile Ile Ala Lys Glu Leu Ile Glu Ile
 130 135 140
 40 Asn
 145

<210> 13
 <211> 435
 45 <212> DNA
 <213> Neisseria meningitidis

<400> 13
 atgggaaatt tcttatatag aggcattagt tgccaacaag atgagcaaaa taatggacag 60
 50 ttaaaaccta aaggtaataa agctgaagtt gcaattcgtt atgatggtaa gtttaaatat 120
 gatggtaaag ctacacatgg tccaagtgtg aagaatgcag tttacgccca tcaaattgaa 180
 acaggtctat atgacggatg ttatatatct acgacaacag acaaggaaat tgccaagaaa 240
 ttgcaacaa gttccggcat cgaaaatggc tatatatatg ttttaaatag ggatttggtt 300
 ggtcaatatt ctatttttga atatggggtt gaacatccag aaaacccaaa tgagaaggaa 360
 55 gtaacaatca gagctgaaga ttgtggctgt attcctgaag aagtgattat tgctaaagag 420
 ttgatagaaa ttaac 435

<210> 14
 <211> 145
 60 <212> PRT
 <213> Neisseria meningitidis

-10-

<400> 14
 Met Gly Asn Phe Leu Tyr Arg Gly Ile Ser Cys Gln Gln Asp Glu Gln
 1 5 10 15
 5 Asn Asn Gly Gln Leu Lys Pro Lys Gly Asn Lys Ala Glu Val Ala Ile
 20 25 30
 10 Arg Tyr Asp Gly Lys Phe Lys Tyr Asp Gly Lys Ala Thr His Gly Pro
 35 40 45
 Ser Val Lys Asn Ala Val Tyr Ala His Gln Ile Glu Thr Gly Leu Tyr
 50 55 60
 15 Asp Gly Cys Tyr Ile Ser Thr Thr Thr Asp Lys Glu Ile Ala Lys Lys
 65 70 75 80
 Phe Ala Thr Ser Ser Gly Ile Glu Asn Gly Tyr Ile Tyr Val Leu Asn
 85 90 95
 20 Arg Asp Leu Phe Gly Gln Tyr Ser Ile Phe Glu Tyr Glu Val Gly His
 100 105 110
 Pro Glu Asn Pro Asn Glu Lys Glu Val Thr Ile Arg Ala Glu Asp Cys
 115 120 125
 25 Gly Cys Ile Pro Glu Glu Val Ile Ile Ala Lys Glu Leu Ile Glu Ile
 130 135 140
 30 Asn
 145
 <210> 15
 <211> 435
 35 <212> DNA
 <213> Neisseria meningitidis
 <400> 15
 atgggaaatt tcttatatag aggcattagt tgccaacaag atgagcaaaa taatggacag 60
 40 ttaaaacctt aaggtataaa agctgaagtt gcaattcggt atgatggtaa gtttaaatat 120
 gatggtaaag ctacacatgg tccaagtgtg aagaatgcag tttagcctca tcaaattgaa 180
 acaggtctat atgacggatg ttatatatct acgacaacag acaaggaaat tgccaagaaa 240
 tttgcaacaa gttccggcat cgaaaatggc tatatatatg ttttaaatag ggatttgttt 300
 ggtcaatatt ctatttttga atatgaggtt ggacatccag aaaaccctaaa tgagaaggaa 360
 45 gtaacaatca gagctgaaga ttgtggctgt attcctgaag aagtgattat tgctaaagag 420
 ttgatagaaa ttaac 435
 <210> 16
 <211> 145
 50 <212> PRT
 <213> Neisseria meningitidis
 <400> 16
 Met Gly Asn Phe Leu Tyr Arg Gly Ile Ser Cys Gln Gln Asp Glu Gln
 1 5 10 15
 Asn Asn Gly Gln Leu Lys Pro Lys Gly Asn Lys Ala Glu Val Ala Ile
 20 25 30
 60 Arg Tyr Asp Gly Lys Phe Lys Tyr Asp Gly Lys Ala Thr His Gly Pro
 35 40 45

-11-

Ser Val Lys Asn Ala Val Tyr Ala His Gln Ile Glu Thr Gly Leu Tyr
50 55 60

5 Asp Gly Cys Tyr Ile Ser Thr Thr Thr Asp Lys Glu Ile Ala Lys Lys
65 70 75 80

Phe Ala Thr Ser Ser Gly Ile Glu Asn Gly Tyr Ile Tyr Val Leu Asn
85 90 95

10 Arg Asp Leu Phe Gly Gln Tyr Ser Ile Phe Glu Tyr Glu Val Glu His
100 105 110

15 Pro Glu Asn Pro Asn Glu Lys Gly Val Thr Ile Arg Ala Glu Asp Cys
115 120 125

Gly Cys Ile Pro Glu Glu Val Ile Ile Ala Lys Glu Leu Ile Glu Ile
130 135 140

20 Asn
145

<210> 17
<211> 435
25 <212> DNA
<213> Neisseria meningitidis

<400> 17
30 atgggaaatt tcttatatag aggcattagt tgccaacaag atgagcaaaa taatggacag 60
ttaaacccta aaggtaataa agctgaagtt gcaattcgtt atgatggtaa gtttaaatat 120
gatggtaaag ctacacatgg tccaagtgtg aagaatgcag tttacgcca tcaaattgaa 180
acaggtctat atgacggatg ttatatatct acgacaacag acaaggaaat tgccaagaaa 240
tttgcaacaa gttccggcat cgaaaatggc tatatatatg ttttaaatag ggatttggtt 300
ggccaatatt ctatttttga atatgaggtt gaacatccag aaaacccaaa tgagaaggga 360
35 gtaacaatca gagctgaaga ttgtggctgt attcctgaag aagtgattat tgctaaagag 420
ttgatagaaa ttaac 435

<210> 18
<211> 45
40 <212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

45 <400> 18
cgcggtatccc atatgggaaa tttcttatat agaggcatta gttgc 45

<210> 19
50 <211> 37
<212> DNA
<213> PCR primer

<400> 19
55 cccgctcgag gtttaatttct atcaactctt tagcaat 37

<210> 20
<211> 45
<212> DNA
60 <213> PCR primer

-12-

	<400> 20		
	cgcggatccc atatgggaaa tttcttatat aaaggcatta gttgc		45
5	<210> 21		
	<211> 33		
	<212> DNA		
	<213> PCR primer		
10	<400> 21		
	atTTTTgaat atgggggttga acatccagaa aac		33
	<210> 22		
	<211> 33		
	<212> DNA		
15	<213> PCR primer		
	<400> 22		
	ttctggatgt tcaaccccat attcaaaaat aga		33
20	<210> 23		
	<211> 27		
	<212> DNA		
	<213> PCR primer		
25	<400> 23		
	tatgaggttg gacatccaga aaaccca		27
	<210> 24		
	<211> 27		
30	<212> DNA		
	<213> PCR primer		
	<400> 24		
35	gttttctgga tgtccaacct catattc		27
	<210> 25		
	<211> 28		
	<212> DNA		
	<213> PCR primer		
40	<400> 25		
	ccaaatgaga aggagtaac aatcagag		28
	<210> 26		
45	<211> 27		
	<212> DNA		
	<213> PCR primer		
	<400> 26		
50	gattgttact cccttctcat ttgggtt		27

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079242 A3(51) International Patent Classification⁷: C07K 14/22,
14/255, 14/30, 14/315, 14/36, C12N 15/31, A61K 39/02,
39/095, 39/112

(21) International Application Number: PCT/IB02/02080

(22) International Filing Date: 28 March 2002 (28.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0108024.1 30 March 2001 (30.03.2001) GB(71) Applicant (for all designated States except US): CHIRON
SRL, [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): MASNANI, Vega
[IT/IT]; Chiron SpA, Via Fiorentina, 1, I-53100 Siena (IT).
PIZZA, Mariagrazia [IT/IT]; Chiron SpA, Via Fiorentina,
1, I-53100 Siena (IT). RAPPUOLI, Rino [IT/IT]; Chiron
SpA, Via Fiorentina, 1, I-53100 Siena (IT).(74) Agents: HALLYBONE, Huw, George et al.; Carpmaels
& Ransford, 43 Bloomsbury Square, London WC1A 2RA
(GB).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Published:

--- with international search report
--- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments(88) Date of publication of the international search report:
6 November 2003For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: ADP-RIBOSYLATING BACTERIAL TOXINS

<i>E.coli</i> LT	KLYRADSRPPD	LYDHARGTQ	YDDGYVSTSLSLRSA	SPHPYEQEVSA LGGI
<i>N.meningitidis</i>	FLYRGISCCQD	VYAHQIETG	YDGCYISTTDKEIA	PENPNEKEVTIRAED
<i>S.coelicolor</i>	TLYRSDSRGPQVVFEFGF	HAKDVQNGQYDVEKYVLVNQ	SPSYVSTSYDHDLY	HKWADQVEVAFPGGI
<i>M.pneumoniae</i>	FVYRVDLRSPE	FFEHLSTN	GRSYFISTSETPTAA	TSFAYQREWFIDGPI
<i>S.typhi</i>	FVYRVDSPTPD		SCSGGSSDSRYIATT	TMRLOREYVSTLSI
<i>S.paratyphi</i>	FVYRVDSPTPD		SCSGGSSDSRYIATT	TMRLOREYVSTLSI
<i>S.pyogenes</i>	VVYRVYETFL		TKHSFMSTTALKNGA	SAVPSEVELLFPRGC

(57) Abstract: ADP-ribosylating toxins from *Neisseria meningitidis*, *Streptomyces coelicolor*, *Mycoplasma pneumoniae*, *Salmonella typhimurium*, *Salmonella paratyphi*, and *Streptococcus pyogenes* are disclosed, together with mutant toxins and uses therefor. There is only a low level of sequence identity between these toxins and toxins such as cholera toxin and *E.coli* heat labile toxin.

WO 02/079242 A3

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/IB 02/02080

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C07K14/22 C07K14/255 C07K14/30 C07K14/315 C07K14/36
 C12N15/31 A61K39/02 A61K39/095 A61K39/112

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TETTELIN H ET AL: "COMPLETE GENOME SEQUENCE OF NEISSERIA MENINGITIDIS SEROGROUP B STRAIN MC58" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 287, 2000, pages 1809-1815, XP000914963 ISSN: 0036-8075 cited in the application	17,18,26
X	-& DATABASE EMBL 'Online!' gene NM81343, 15 March 2000 (2000-03-15) "N. meningitidis serogroup B strain MC58 section 124 of 206 of the complete genome" retrieved from EBI Database accession no. AE002482 XP002239404 related to SEQ ID No. 2; 100% identity in 435 nt overlap from pos. 9519-9953 abstract -/--	17,18,26



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

25 April 2003

Date of mailing of the international search report

31.02.03

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
 Fax: (+31-70) 340-3016

Authorized officer

Alt, G

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 02/02080

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	cds 9519..9956; hypothetical protein	1
A	-& DATABASE SWALL 'Online! 15 October 2000 (2000-10-15) "Hypothetical protein NMB 1343" retrieved from EBI Database accession no. Q9JZ10 XP002239405 related to SEQ ID No. 1; 100% identity in 145nt overlap from pos. 1-145 abstract	1
X	ALA'ALDEEN D A A ET AL: "CLONING, SEQUENCING, CHARACTERISATION AND IMPLICATIONS FOR VACCINE DESIGN OF THE NOVEL DIHYDROLIPOYL ACETYLTRANSFERASE OF NEISSERIA MENINGITIDIS" JOURNAL OF MEDICAL MICROBIOLOGY, HARLOW, GB, vol. 45, December 1996 (1996-12), pages 419-432, XP009009026 ISSN: 0022-2615 page 424, second column, lines 1-3 of second paragraph and lines 1-8 of third paragraph	17,18,26
X	-& DATABASE EMBL 'Online! 25 November 1994 (1994-11-25) ALA'ALDEEN, D.: "N. meningitidis pdh gene cluster" retrieved from EBI Database accession no. X82637 XP002239406 related to SEQ ID No. 2; 100% identity in 435 nt overlap from pos. 4708-5142 abstract	17,18,26
X	DATABASE EMBL 'Online! 8 March 1994 (1994-03-08) GUILLEN, G.: "N. meningitidis outer membrane protein P64k (pm-6) gene" retrieved from EBI Database accession no. X77920 XP002239407 related to SEQ ID No. 2; 99.7 % identity in 435 nt overlap from pos. 170-603 abstract	17,18,26

-/--

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 02/02080

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL 'Online! 4 December 2000 (2000-12-04) FRAZER, C.M. ET AL.: "N. meningitidis partial DNA sequence gnm_35 SEQ ID No. 35" retrieved from EBI Database accession no. AAA81487 XP002239408 related to SEQ ID No. 2; 100% identity in 435 nt overlap from pos. 24115-24549 abstract & WO 00 22430 A (CHIRON CORP.) 20 April 2000 (2000-04-20)</p>	17,18,26
X	<p>DATABASE EMBL 'Online! 24 November 2000 (2000-11-24) PIZZA, M. ET AL.: "Neisseria genomic sequences; Sequence 111" retrieved from EBI Database accession no. AX044032 XP002239409 related to SEQ ID No. 2; 100% identity in 435 nt overlap from pos. 163680-164114 abstract & WO 00 66791 A (CHIRON CORP.; THE INSTITUTE FOR GENOMIC RESEARCH) 9 November 2000 (2000-11-09)</p>	17,18,26
X	<p>DATABASE EMBL 'Online! 25 March 1991 (1991-03-25) DAMS, E. ET AL.: "Cholera toxin A protein (CTA)" retrieved from EBI Database accession no. CAA41592 XP002239410 cited in the application abstract</p>	2,4-6,9
X	<p>DATABASE EMBL 'Online! MASON, H. S. AND ARNTZEN, C.J.: "Plant-optimized mutant V.cholera toxin subunit K63" retrieved from EBI Database accession no. AAY966654 XP002239411 100% identity in 258 nt overlap with sequence of accession number CAA41592 abstract</p>	10

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB 02/02080

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GIUDICE DEL G ET AL: "GENETICALLY DERIVED TOXOIDS FOR USE AS VACCINES AND ADJUVANTS" VACCINE, BUTTERWORTH SCIENTIFIC. GUILDFORD, GB, vol. 17, no. SUPPL 2, 1 October 1999 (1999-10-01), pages S44-S52, XP000996148 ISSN: 0264-410X the whole document	12-15, 22-25
A	KRUEGER, K. AND BARBIERI, J.T.: "The family of bacterial ADP-ribosylating exotoxins" CLINICAL MICROBIOLOGY REVIEWS, vol. 8, no. 1, January 1995 (1995-01), pages 34-47, XP002239403 the whole document	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 02/02080

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 25 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

1-26
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-26

Protein related to SEQ ID No. 1; Mutants of that protein, specifically SEQ ID Nos. 10, 12, 14 and 16; DNA encoding the protein, specifically SEQ ID No. 2, 11, 13, 15, 17; vectors and host cells comprising the above nucleic acids; antibodies; compositions related to the above protein or nucleic acid; use of the protein and method of treatment

2. Claims: 1-7, 9-26

as subject 1, but related to SEQ ID No. 3,

3. Claims: 1-7, 9-26

as subject 1, but related to SEQ ID No. 4

4. Claims: 1-7, 9-26

as subject 1, but related to SEQ ID No. 5

5. Claims: 1-7, 9-26

as for subject 1, but related to SEQ ID No. 6

6. Claims: 1-7, 9-26

as for subject 1, but related to SEQ ID No. 7

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 02/02080

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0022430 A	20-04-2000	AU 1202200 A	01-05-2000
		AU 761780 B	12-06-2003
		AU 3967799 A	23-11-1999
		BR 9914374 A	17-09-2002
		CA 2330838 A	11-11-1999
		CA 2346713 A	20-04-2000
		CN 1373806 T	09-10-2002
		CN 1338005 T	27-02-2002
		EP 1093517 A	25-04-2001
		EP 1144998 A	17-10-2001
		AU 3249200 A	17-11-2000
		BR 0010361 A	10-06-2003
		CA 2371032 A	09-11-2000
		CN 1359426 T	17-07-2002
		EP 1185691 A	13-03-2002
		WO 0066791 A	09-11-2000
WO 0066791 A	09-11-2000	AU 3249200 A	17-11-2000
		AU 3947801 A	12-09-2001
		AU 3948801 A	12-09-2001
		BR 0010361 A	10-06-2003
		CA 2371032 A	09-11-2000
		CA 2400562 A	07-09-2001
		CA 2400570 A	07-09-2001
		CN 1359426 T	17-07-2002
		CN 1426471 T	25-06-2003
		CN 1426473 T	25-06-2003
		EP 1185691 A	13-03-2002
		EP 1261723 A	04-12-2002
		EP 1259627 A	27-11-2002
		WO 0164920 A	07-09-2001
		WO 0164922 A	07-09-2001
		AU 1202200 A	01-05-2000
		BR 9914374 A	17-09-2002
		CA 2346713 A	20-04-2000
		CN 1338005 T	27-02-2002
		EP 1144998 A	17-10-2001
		WO 0022430 A	20-04-2000